

Routing under Balance

Alina Ene*

Gary Miller†

Jakub Pachocki‡

Aaron Sidford§

Abstract

We introduce the notion of *balance* for directed graphs: a weighted directed graph is α -balanced if for every cut $S \subseteq V$, the total weight of edges going from S to $V \setminus S$ is within factor α of the total weight of edges going from $V \setminus S$ to S . Several important families of graphs are nearly balanced, in particular, Eulerian graphs (with $\alpha = 1$) and residual graphs of $(1 + \epsilon)$ -approximate undirected maximum flows (with $\alpha = \mathcal{O}(1/\epsilon)$).

We use the notion of balance to give a more fine-grained understanding of several well-studied routing questions that are considerably harder in directed graphs. We first revisit oblivious routings in directed graphs. Our main algorithmic result is an oblivious routing scheme for *single-source* instances that achieve an $\mathcal{O}(\alpha \cdot \log^3 n / \log \log n)$ competitive ratio. In the process, we make several technical contributions which may be of independent interest. In particular, we give an efficient algorithm for computing *low-radius decompositions* of directed graphs parameterized by balance. We also define and construct *low-stretch arborescences*, a generalization of low-stretch spanning trees to directed graphs.

On the negative side, we present new lower bounds for oblivious routing problems on directed graphs. We show that the competitive ratio of oblivious routing algorithms for directed graphs is $\Omega(n)$ in general; this result improves upon the long-standing best known lower bound of $\Omega(\sqrt{n})$ [HKRL07]. We also show that our restriction to single-source instances is necessary by showing an $\Omega(\sqrt{n})$ lower bound for multiple-source oblivious routing in Eulerian graphs.

We also study the maximum flow problem in balanced directed graphs with arbitrary capacities. We develop an efficient algorithm that finds an $(1 + \epsilon)$ -approximate maximum flows in α -balanced graphs in time $\tilde{\mathcal{O}}(m\alpha^2/\epsilon^2)$. We show that, using our approximate maximum flow algorithm, we can efficiently determine whether a given directed graph is α -balanced. Additionally, we give an application to the directed sparsest cut problem.

*Department of Computer Science and DIMAP, University of Warwick, A.Ene@dcs.warwick.ac.uk.

†Carnegie Mellon University, glmiller@cs.cmu.edu.

‡Carnegie Mellon University, pachocki@cs.cmu.edu.

§Microsoft Research New England, asid@microsoft.com.

1 Introduction

In this paper, we study several fundamental routing questions in *directed graphs* that are *nearly Eulerian*. We introduce the notion of *balance* for directed graphs that quantifies how far away a graph is from being Eulerian¹: a weighted directed graph is α -balanced if for every cut $S \subseteq V$, the total weight of edges going from S to $V \setminus S$ is within factor α of the total weight of edges going from $V \setminus S$ to S . Several important families of graphs are nearly balanced, in particular, Eulerian graphs (with $\alpha = 1$) and residual graphs of $(1 + \epsilon)$ -approximate undirected maximum flows (with $\alpha = \mathcal{O}(1/\epsilon)$).

We use the notion of balance to give a more fine-grained understanding of several well-studied routing questions that are considerably harder in directed graphs. The first question that we address is that of designing *oblivious routing* schemes for directed graphs. Oblivious routing schemes were introduced in the seminal work of Räcke [Räc02]. They are motivated by practical applications in routing traffic in massive networks such as the Internet, where it is necessary to route each request independently of the other requests and the current traffic in the network. Oblivious routing schemes were developed in a sequence of works [Räc02, ACF⁺03, BKR03, HKLR05, HKLR06, HKRL07, Räc08, ER09]. In particular, if the graph is undirected, there exist oblivious routing schemes that achieve competitive ratio $\mathcal{O}(\log n)$ [Räc08], where n is the number of nodes, and this result is optimal [BL99, MMVW97, MMVW97]. In contrast, Hajiaghayi *et al.* [HKRL07] show a strong lower bound of $\Omega(\sqrt{n})$ on the competitive ratio of routing obliviously in directed graphs. This lower bound holds even for *single-source* instances of bounded degree graphs, as well as for instances with symmetric demands.

In this paper, we revisit oblivious routing in directed graphs, and we show that balanced graphs bridge the gap between directed and undirected graphs (see Section 3). Our main algorithmic result is an oblivious routing scheme for *single-source* instances that achieve an $\mathcal{O}(\alpha \cdot \log^3 n / \log \log n)$ competitive ratio. In the process, we make several technical contributions which may be of independent interest. In particular, we give an efficient algorithm for computing *low-radius decompositions* of directed graphs parameterized by balance. We also define and construct *low-stretch arborescences*, a new concept generalizing low-stretch spanning trees to directed graphs. Given the far-reaching implications of low-diameter decompositions and low-stretch spanning trees, we hope that our techniques may find other applications.

Our result is a generalization to directed graphs of Räcke’s influential work [Räc08] that established a remarkable connection between oblivious routing in undirected graphs and metric embeddings into trees.

On the negative side, we present new lower bounds for oblivious routing problems on directed graphs. We show that the competitive ratio of oblivious routing algorithms for directed graphs has to be $\Omega(n)$ in general; this result improves upon the long-standing best known lower bound of $\Omega(\sqrt{n})$ [HKRL07]. We also show that the restriction to single-source instances is necessary by showing an $\Omega(\sqrt{n})$ lower bound for multiple-source oblivious routing in Eulerian graphs.

The second question that we study is that of finding an *approximate maximum flow* in balanced graphs. The maximum flow problem has received considerable attention in recent years, leading to several breakthrough results. This line of work has led to the development of almost linear time algorithms for approximate maximum flows in undirected graphs [KLOS14, She13] and the

¹A directed graph is Eulerian if, for each vertex, the total weight of its incoming edges is equal to the total weight of its outgoing edges. An equivalent definition is that for each cut $S \subseteq V$, the total weight of edges from S to $V \setminus S$ is equal to the total weight of edges from $V \setminus S$ to S .

subsequent improvement of [Pen14, RST14]. In contrast, progress on directed graphs has been comparatively more modest, and the only improvements are the breakthrough results of Madry, yielding an $\tilde{O}(m^{10/7})$ -time algorithm for *unit-capacity* directed graphs with m edges [Mad13] and of Lee and Sidford, obtaining a running time of $\tilde{O}(m\sqrt{n})$ for arbitrary directed graphs [LS13]. These improve over the long-standing best running time of $\tilde{O}(m \min(\sqrt{m}, n^{2/3}))$ given by Goldberg and Rao [GR98].

In this paper, we study the maximum flow problem in balanced directed graphs with arbitrary capacities (see Section 5). We develop an efficient algorithm that finds an $(1 + \epsilon)$ -approximate maximum flows in α -balanced graphs in time $\tilde{O}(m\alpha^2/\epsilon^2)$. Our algorithm builds on the work of Sherman [She13] and it can be viewed as an analogue of his result for directed graphs. The running time of our algorithm degrades gracefully with the imbalance of the graph and thus it suggests that balanced graphs provide a meaningful bridge between undirected and directed graphs.

We show that, using our approximate maximum flow algorithm, we can efficiently determine whether a given directed graph is α -balanced (see Section 5.2). Additionally, we give an application to the directed sparsest cut problem (see Section 5.3).

1.1 Related Work

Oblivious Routing. Oblivious routing schemes are well-studied and several results are known; we refer the reader to [Räc09] for a comprehensive survey of results for undirected graphs. As mentioned previously, in edge-weighted undirected graphs one can achieve a competitive ratio of $\mathcal{O}(\log n)$ [Räc08], and it is the best possible [BL99, MMVW97, MMVW97]. Hajiaghayi *et al.* [HKRL07] studied oblivious routing schemes in node-weighted undirected graphs and directed graphs. Their work gives an $\Omega(\sqrt{n})$ lower bound on the competitive ratio for both node-capacitated undirected graphs and directed graphs. They also show that these lower bounds still hold in more restricted settings, such as single-source instances. On the positive side, they give oblivious routing scheme with competitive ratios of $\mathcal{O}(\sqrt{n} \log n)$ for single-source instances in bounded-degree directed graphs, and $\mathcal{O}(\sqrt{kn}^{1/4} \log n)$ for general instances in directed graphs, where k is the number of commodities and in the worst case $k = \Theta(n^2)$.

Maximum s - t Flows. The maximum flow problem is one of the most central problems in combinatorial optimization and has been studied extensively over the past several decades. Until recently, most approaches have been based on combinatorial methods such as augmenting paths, blocking flows, push-relabel, etc. This line of work culminated in the seminal algorithm of Goldberg and Rao [GR98] that computes a maximum flow in time $O(\min(n^{2/3}, m^{1/2}) \log(n^2/m) \log U)$ in directed graphs with integer weights that are at most U .

Over the past decade, a new approach emerged based on techniques drawn from several areas such as continuous optimization, numerical linear algebra, and spectral graph theory. These approaches led to a nearly-linear time algorithm for approximate maximum flows in *undirected graphs* [She13, KLOS14, Pen14], an $\tilde{O}(m^{10/7})$ -time algorithm for maximum flows in *unit-capacity directed graphs* [Mad13] and an $\tilde{O}(m\sqrt{n})$ -time algorithm for arbitrary directed graphs [LS13].

1.2 Organization

The rest of this paper is organized as follows. In Section 2, we give an overview of our main results and introduce the definitions and notation we use throughout the paper. In Section 3, we give our oblivious routing scheme for single-source instances. In Section 4, we state our lower bounds for oblivious routing. In Section 5 we give our approximate maximum flow algorithm and applications. Many proofs are deferred to the Appendix.

2 Overview

2.1 Basic Definitions

We study directed graphs $G = (V, E, w, l)$ with edge set $E \subseteq V \times V$, edge weights $w : E \rightarrow \mathbb{R}_+$ and edge lengths $l : E \rightarrow \mathbb{R}_+$. Throughout this paper, we assume that G is strongly connected. In several applications we deal with graphs without weights or lengths. For graphs with edge lengths, we let $d(u, v)$ denote the shortest path distance from u to v .

We associate the following matrices with the graph G . The matrix of edge weights is defined as $\mathbf{C} \stackrel{\text{def}}{=} \text{diag}(w)$ and the vertex-edge incidence matrix $\mathbf{B} \in \mathbb{R}^{V \times E}$ is defined as $\mathbf{B}_{s,(u,v)} \stackrel{\text{def}}{=} -1$ if $s = u$, 1 if $s = v$ and 0 otherwise. We are interested in finding flows that route demands with low congestion. The congestion incurred by a flow f is $\|\mathbf{C}^{-1}f\|_\infty$, and we say f routes demands b if $\mathbf{B}f = b$. The problem of finding a minimum congestion flow for a given demand vector, and its dual, the maximum congested cut, can be formulated as follows:

$$\begin{aligned} \min_f \quad & \|\mathbf{C}^{-1}f\|_\infty \quad \text{s.t.} \quad \mathbf{B}f = d, f \geq 0. \\ \max_v \quad & b^\top v \quad \text{s.t.} \quad \|\mathbf{C} \max(\mathbf{B}^\top v, 0)\|_1 \leq 1. \end{aligned}$$

We let OPT_b denote the optimum value of these problems. Throughout the paper, we let $b_S = \sum_{u \in S} b_u$ and $w(S, T)$ denote the total weight of edges from S to T . It is well-known that for the second problem, one of the threshold cuts with respect to v achieves $b_S/w(S, V - S) \geq b^\top v$.

2.2 Balance

We parameterize strongly connected directed graphs by their *imbalance*:

Definition 2.1 (Imbalance) Let $G = (V, E, w)$ be a strongly connected directed graph. We define its imbalance, $\text{bal}(G)$, as the minimum α such that $w(S, V \setminus S) \leq \alpha \cdot w(V \setminus S, S)$ for every $S \subseteq V$.

Two canonical families of balanced graphs are Eulerian graphs. and residual graphs of approximate undirected maximum flows.

Fact 2.2 A strongly connected directed graph G is Eulerian if and only if $\text{bal}(G) = 1$. If G is the residual graph of a $(1 + \epsilon)$ -approximate undirected maximum flow, then $\text{bal}(G) = \mathcal{O}(\epsilon^{-1})$.

Theorem 2.3 (Equivalent definitions of balance) Let $G = (V, E, w)$ be a directed graph. The following statements are equivalent:

1. $\text{bal}(G) \leq \alpha$.
2. There exists a circulation f on G with all edge congestions in $[1, \alpha]$.
3. Let $d = \mathbf{B}\vec{1}$ be the residual degrees in G . Then $-d$ can be routed with congestion $\alpha - 1$.

2.3 Oblivious Routing Schemes

An oblivious routing scheme is a linear operator that, for each source-destination pair $(s, t) \in V \times V$, specifies how to route one unit of flow from s to t independently of the other pairs. Given a demand vector $\vec{d} : D \rightarrow \mathbb{R}_+$ on a set $D \subseteq V \times V$ of source-sink pairs, one can produce a multi-commodity flow that meets these demands by routing each demand pair using the (pre-specified) operator, independently of the other demands. The *competitive ratio* of an oblivious routing scheme is the worst ratio among all possible demand vectors between the congestion of the multi-commodity flow given by the scheme and the congestion of the minimum congestion multi-commodity flow for the

given demand vector.

Our main positive result concerning oblivious routings, given in Section 3, is the existence of good single-source oblivious routings for balanced graphs. A single-source oblivious routing with source $s \in V$ has $D = \{s\} \times V$.

Theorem 2.4 (Single Source Oblivious Routings) *Every strongly connected graph G admits a single-source oblivious routing, from any source, with competitive ratio $\mathcal{O}(\text{bal}(G) \cdot \log^3 n / \log \log n)$.*

We achieve this result by generalizing an algorithm for undirected graphs given by Racke [Räc08]. The core difficulty that we need overcome is to find a good way to cluster the vertices of a directed balanced graph. We define the *radius* of a cluster $C \subseteq V$ as $\min_{u \in C} \max_{v \in C} d(u, v)$. The *volume* $\text{vol}(G)$ of G is defined as $\text{vol}(G) \stackrel{\text{def}}{=} \sum_{e \in E} l(e)w(e)$. Our clustering algorithm is presented in Section 3.1, and its guarantees can be formalized as follows:

Theorem 2.5 (Balanced Graph Clustering) *Let $G = (V, E, w, l)$ be a directed graph. Then for every $r > 0$, V can be partitioned into clusters such that every cluster has radius at most r , and the total weight of edges going between different clusters is $\mathcal{O}(\text{bal}(G)\text{vol}(G) \log n/r)$. Moreover, such a partition can be found in expected linear time.*

The guarantees of Theorem 2.5 for undirected graphs match those given by prior work [Awe85, AKPW95, Bar96, MPX13]. Extending the statement to directed graphs is nontrivial, as it requires making the notion of cluster radii directed.

In Section 4 we give a new lower bound for all-pairs oblivious routings in directed graphs.

Theorem 2.6 *No oblivious routing algorithm for directed graphs can guarantee competitive ratio better than $\Omega(n)$.*

We also show that restricting ourselves to single-source oblivious routings is necessary to achieve a small competitive ratio even when $\text{bal}(G) = 1$.

Theorem 2.7 *No oblivious routing algorithm for Eulerian graphs can guarantee competitive ratio better than $\Omega(\sqrt{n})$.*

2.4 Maximum Flows

Finally, we consider the maximum s - t flow problem in directed graphs parameterized by balance. Given a source s and a destination t , the maximum s - t flow problem asks us to find a flow f that routes as much flow as possible from s to t while sending at most w_e units of flow along each edge e . In Section 5 we show the following result.

Theorem 2.8 (Approximate Maximum Flow) *Given a strongly connected directed graph G , a source s , and a sink t there is an algorithm that finds a $(1 + \epsilon)$ -approximate maximum s - t flow and a $(1 - \epsilon)$ -approximate minimum s - t cut in G in time $\tilde{\mathcal{O}}(m \cdot \text{bal}(G)^2 / \epsilon^2)$.*

To achieve quadratic dependency on ϵ , in Section 5.4 we provide a general analysis of gradient descent for composite function minimization under non-Euclidean norms.

We also show applications of this result to computing the sparsest cut (Section 5.3) and we prove the following result on computing the imbalance of a graph (Section 5.2).

Lemma 2.9 *There is an algorithm that either certifies that $\text{bal}(G) \leq \alpha$ or shows that $\text{bal}(G) > (1 - \epsilon)\alpha$ in time $\tilde{\mathcal{O}}(m\alpha^2/\epsilon^2)$.*

3 Oblivious Routing on Balanced Graphs

3.1 Low-radius Decompositions

Our algorithm for clustering directed graphs, presented in Figure 1, is based on the scheme given by Miller, Peng and Xu [MPX13]. We first pick a start time x_v for every vertex v from an exponential distribution, and then explore the graph, starting the search from v at time x_v and proceeding at unit speed. Each vertex u is assigned to the vertex v that reached it first.

$(V_1, V_2, \dots) = \text{CLUSTER-DIRECTED}(G, r)$, where $G = (V, E, l)$ is a directed graph and $r > 0$.

1. Set $\beta := \log n / (10r)$.
2. For every vertex $v \in V$ pick $x_v \sim \text{Exp}(\beta)$.²
3. For each vertex $u \in V$, assign u to the cluster rooted at the vertex $v \in V$ which minimizes $-x_v + d(v, u)$.
4. If any of the clusters has radius greater than r , return to step 2. Otherwise, return the clusters.

Figure 1: The low-radius decomposition algorithm for directed graphs.

Our goal is to show that this procedure *cuts* few edges, i.e. assigns the endpoints of few edges to different clusters. The original analysis of [MPX13] shows that for undirected graphs, this approach guarantees cutting each edge e with low probability, namely $\mathcal{O}(l(e) \log n / r)$. It turns out that even in the case of unweighted Eulerian graphs such a guarantee no longer holds; there may exist edges that are cut with very high probability. Consider for instance (Figure 2) a directed cycle of length 3^k , with an undirected star of 2^{k^2} leaves attached to one of its vertices, v . Set $r := 2^k$. Let u be the vertex preceding v on the cycle. It is now easy to verify by calculation that the edge (u, v) is cut with probability arbitrarily close to 1 for a large enough k . With high probability, v will be contained in a cluster rooted at one of the 2^{k^2} leaves attached to it; also with high probability, no such cluster will contain u .

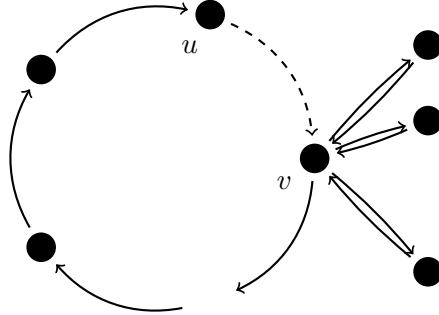


Figure 2: An unweighted Eulerian graph where a particular edge is very likely to be cut by the scheme of [MPX13].

This issue requires us to find a new way to guarantee that the total weight of cut edges is low. Our key idea is to show that, for any fixed *cycle*, the expected number of edges in the cycle that are cut is small. The desired guarantees then follow by noting that any graph G can be approximated up to a factor $\text{bal}(G)$ by a sum of cycles (Theorem 2.3).

² $\text{Exp}(\beta)$ is the exponential distribution with parameter β , with p.d.f. $f(x) = \beta e^{-\beta x}$ on $x \geq 0$.

Lemma 3.1 *Let \mathcal{P} be the partition returned by $\text{CLUSTER-DIRECTED}(G, r)$. For any simple cycle C in G , the expected number of edges in C that go between different clusters in \mathcal{P} is an $\mathcal{O}(\log n/r)$ fraction of the length of C .*

As the above example demonstrates, we cannot base the proof of Lemma 3.1 on the *location* of the cuts, as it might depend strongly on the input graph. However, we can prove that, intuitively, cuts occur *infrequently* as the graph is explored. This is the crucial idea of the proof: we analyze the occurrence of cuts over time rather than bounding the probabilities of particular cuts. Then we use the fact that a cycle of length L is fully explored within L time steps after the time it is visited for the first time. The analysis is presented in Appendix B.

3.2 Low-stretch Arborescences

Let G be a directed graph and let s be a vertex in G . We say that a directed graph T is an *arborescence* rooted at s for every vertex v , there is a unique directed path in T from s to v . In this section, we define and construct *low-stretch arborescences*, which are a key intermediate step between low-radius decompositions and oblivious routings.

Definition 3.2 *Let $G = (V, E, w, l)$ be a directed graph. We define the stretch of an edge $(u, v) \in E$ with respect to an arborescence T on the vertex set V as $w(u, v) \cdot d_T(u, v)$, where $d_T(u, v)$ is the distance between u and v in the **undirected** tree corresponding to T .*

Following the notation of [Räc08], we define the *load*, $\text{load}_T(e)$, of an edge $e \in T$ as the sum of the weights of edges $(u, v) \in E(G)$ such that e is on the path between u and v in the *undirected* tree corresponding to T . Note that the total load of the edges in T is equal to the total stretch of the edges in G .

In order to construct low-stretch arborescences, we will recursively cluster V using the algorithm from the previous section. The algorithm FIND-ARBORESCENCE is defined and analyzed in Appendix C. It is similar to the scheme given by Bartal [Bar96]. One major difficulty is that the clusters returned by CLUSTER-DIRECTED may be very imbalanced; in particular, they need not be strongly connected. In order to resolve this issue, we introduce the notion of *additive imbalance* and prove that our clustering algorithms still give good guarantees for graphs with low additive imbalance.

Theorem 3.3 *Let $G = (V, E, w, l)$ be a strongly connected directed graph. Let $s \in V$. Let $T = \text{FIND-ARBORESCENCE}(G, s)$. Then:*

- T has vertex set V and is rooted at s ,
- every arc (u, v) in T can be mapped to a path from u to v in G of equal length, and
- the expected total stretch of G with respect to T is $\mathcal{O}(\text{bal}(G)\text{vol}(G)\log^3 n / \log \log n)$.

Moreover, the algorithm works in expected $\mathcal{O}(m \log n)$ time.

3.3 Constructing the Routing

Given an algorithm for constructing low-stretch arborescences, we can use it to compute a good oblivious routings using the approach proposed by [Räc08]. The oblivious routing that we construct for a given source s will be a convex combination of arborescences rooted at s , with the flow for demand (s, u) being defined as the convex combination of the corresponding paths. The algorithm is given in Figure 3.

The key idea we employ to extend the analysis of the algorithm to a directed graph G is to prove that the routing scheme we construct is competitive even for the *undirected* graph underlying G .

$((T_1, \lambda_1), \dots, (T_k, \lambda_k)) = \text{FIND-ROUTING}(G, s)$ where $G = (V, E, w)$ is a strongly connected directed graph and $s \in V$.

1. Set $k := 0$ and $p_e^{(0)} := 1$ for all $e \in E$.
2. While $\sum_{i=1}^k \lambda_i < 1$:
 - (a) $k := k + 1$.
 - (b) Let $G_k = (V, E, l_k)$ be a copy G with edge lengths

$$l_k(e) := p_e^{(k-1)} / \left(w(e) \sum_{e'} p_{e'}^{(k-1)} \right).$$

- (c) $T_k := \text{FIND-ARBORESCENCE}(G, s)$ (pick the minimum-stretch arborescence out of $\mathcal{O}(\log n)$ runs).
- (d) $\ell_k := \max_e \{\text{load}_{T_k}(e)/w(e)\}$.
- (e) $\lambda_k := \min \left(1/\ell_k, 1 - \sum_{i=1}^{k-1} \lambda_i \right)$.
- (f) For all edges e set:

$$p_e^{(k)} := p_e^{(k-1)} \cdot \exp(\lambda_k \cdot \text{load}_{T_k}(e)/w(e)).$$

3. Return $((T_1, \lambda_1), \dots, (T_k, \lambda_k))$.

Figure 3: The algorithm for finding single-source oblivious routings on balanced graphs (adapted from [Räc08]).

Lemma 3.4 ([Räc08], adapted) *Let G be a strongly connected directed graph and s be a vertex in G . Let $((T_1, \lambda_1), \dots, (T_k, \lambda_k)) := \text{FIND-ROUTING}(G, s)$. Then with high probability $((T'_1, \lambda_1), \dots, (T'_k, \lambda_k))$ is an $\mathcal{O}(\text{bal}(G) \log^3 n / \log \log n)$ -competitive oblivious routing for G' , where T'_1, \dots, T'_k, G' are the undirected counterparts of T_1, \dots, T_k and G , respectively, that we obtain by ignoring the directions.*

In order to finish the analysis, we only need to note that $((T_1, \lambda_1), \dots, (T_k, \lambda_k))$ is an oblivious routing for G .

Proof of Theorem 2.4: We prove that for any s , the output of $\text{FIND-ROUTING}(G, s)$ satisfies the criteria stated in the theorem statement. It follows from Lemma 3.4 that with high probability, $((T'_1, \lambda_1), \dots, (T'_k, \lambda_k))$ is an $\mathcal{O}(\text{bal}(G) \log^3 n / \log \log n)$ -competitive oblivious routing for G' , where T'_1, \dots, T'_k, G' are undirected counterparts of T_1, \dots, T_k, G , respectively. In particular, it is also an $\mathcal{O}(\text{bal}(G) \log^3 n / \log \log n)$ -competitive oblivious routing from s . Now it is enough to observe that since T_1, \dots, T_k are directed away from s , $((T_1, \lambda_1), \dots, (T_k, \lambda_k))$ is an oblivious routing from s in G . Since it is $\mathcal{O}(\text{bal}(G) \log^3 n / \log \log n)$ -competitive in G' , it must also be $\mathcal{O}(\text{bal}(G) \log^3 n / \log \log n)$ -competitive in G . ■

4 Lower Bounds

We prove new lower bounds for oblivious routings in directed graphs. The constructions and proofs are given in Appendix E.

Theorem 2.6 *No oblivious routing algorithm for directed graphs can guarantee competitive ratio better than $\Omega(n)$.*

Theorem 2.7 *No oblivious routing algorithm for Eulerian graphs can guarantee competitive ratio*

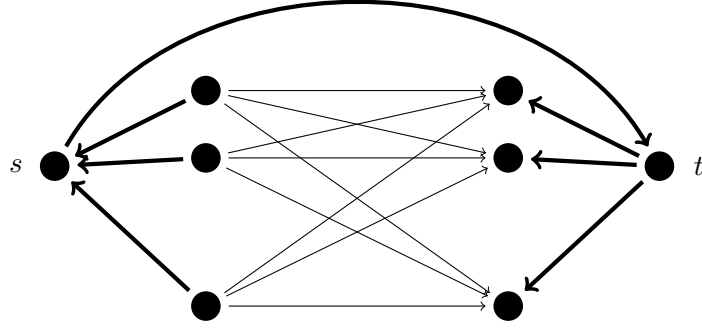


Figure 4: The example from Theorem 2.6. The thick edges have weight n , the other edges have weight 1. Any oblivious routing must put too much flow on the edge (s, t) when routing between the vertices of the biclique.

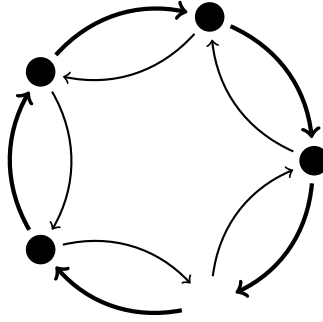


Figure 5: The example from Theorem 2.7. The thick edges have weight \sqrt{n} , the other edges have weight 1. Any oblivious routing must put too much flow on the outer cycle when routing between consecutive vertices of the inner cycle.

better than $\Omega(\sqrt{n})$.

5 Maximum Flow and Applications

5.1 Directed Maximum Flow

In this subsection we show how to efficiently compute an $(1 + \epsilon)$ -approximate maximum flow in directed graphs given a good *congestion-approximator*.

Definition 5.1 An α -congestion-approximator for G is a matrix \mathbf{R} such that for any demand vector b , $\|\mathbf{R}b\|_\infty \leq OPT_b \leq \alpha \|\mathbf{R}b\|_\infty$.

Since $\|\mathbf{R}b\|_\infty = \|-\mathbf{R}b\|_\infty$, only well-balanced graphs admit good congestion approximators:

Fact 5.2 If G admits an α -congestion approximator, $\text{bal}(G) \leq \alpha$.

For undirected graphs, $\tilde{O}(1)$ -congestion-approximators can be computed in nearly linear time [Mad10, She13, KLOS14, Pen14]. This implies that for directed G we can compute $\tilde{O}(\text{bal}(G))$ -congestion-approximators in nearly linear time by the following fact:

Fact 5.3 Let G be a directed graph and G' be its undirected copy. Then for any demand vector b $OPT_b(G') \leq OPT_b(G) \leq (1 + \text{bal}(G))OPT_b(G')$.

Our main result is the following:

Theorem 5.4 *Let G be a directed graph. Given an α -congestion-approximator \mathbf{R} , we can compute an $(1 + \epsilon)$ -approximate maximum flow and minimum congested cut for any demand vector in time $\tilde{\mathcal{O}}(m\alpha^2/\epsilon^2)$, assuming multiplication by \mathbf{R} and \mathbf{R}^\top can be done in $\tilde{\mathcal{O}}(m)$ time.*

Our algorithm is based very heavily on the approach for undirected graphs given by Sherman [She13]. The main difference is the implementation of the key optimization procedure, presented in Figure 6. Due to space constraints, in this section we only outline the main changes needed to extend the algorithm of [She13] to balanced graphs.

Let G be a directed graph and b be a demand vector. Assume we are given an α -congestion-approximator \mathbf{R} . Let $\text{lmax}(x) \stackrel{\text{def}}{=} \ln \sum_i (e^{x_i} + e^{-x_i})$ and define

$$\begin{aligned}\mu(f) &\stackrel{\text{def}}{=} \text{lmax}(2\alpha\mathbf{R}(b - \mathbf{B}f)) \\ \phi(f) &\stackrel{\text{def}}{=} \|\mathbf{C}^{-1}f\|_\infty + \mu(f)\end{aligned}$$

$(f, v) = \text{ALMOST-ROUTE-DIRECTED}(b, \epsilon, f_0)$

1. Initialize $f := f_0, \delta := \frac{\epsilon}{10\alpha^2}$.
2. Scale f and b so that $\|\mathbf{C}^{-1}f\|_\infty + 2\alpha\|\mathbf{R}(b - \mathbf{B}f)\|_\infty = 20\epsilon^{-1} \ln n$.
3. Repeat while any of the following conditions is satisfied:
 - (a) if $\phi(f) < 16\epsilon^{-1} \ln n$, scale f and b up by $17/16$ and restart step 3.
 - (b) let s be $w(e)$ on the coordinates e where $\nabla\mu(f)$ is negative and 0 elsewhere. If $-\nabla\mu(f)^\top s > 1 + \frac{\epsilon}{4}$, set $f := f + \delta s$ and restart step 3.
 - (c) if $\|\mathbf{C}^{-1}f\|_\infty + \nabla\mu(f)^\top f > \frac{\epsilon}{4}$, set $f := f - \delta f$ and restart step 3.
4. Set $x := 2\alpha\mathbf{R}(b - \mathbf{B}f)$.
5. Set $p := \nabla\text{lmax}(x)$.
6. Set $v := \mathbf{R}^\top p$.

Figure 6: The algorithm for computing the maximum flow and minimum congested cut.

Lemma 5.5 *After $\text{ALMOST-ROUTE-DIRECTED}(b, \epsilon, f_0)$ terminates, we have*

$$\phi(f) \leq (1 + \epsilon) \frac{b^\top v}{\|\mathbf{C} \max(\mathbf{B}^\top v, 0)\|_1},$$

assuming $\epsilon \leq 1/2$.

Lemma 5.6 *$\text{ALMOST-ROUTE-DIRECTED}(b, \epsilon, f_0)$ terminates within $\tilde{\mathcal{O}}(\log(1 + \epsilon_0)\alpha^2/\epsilon^3)$ iterations, where $\epsilon_0 = \max(\phi(f_0)/OPT_b - 1, \epsilon)$, assuming $\epsilon \leq 1/2$.*

Note that Lemma 5.5 implies that v is a potential vector for a $(1 + \epsilon)$ -approximate minimum congested cut. In order to recover the corresponding flow, we can employ the recursion described in [She13]. The only additional component necessary for directed graphs is an $\mathcal{O}(\text{poly}(n, \alpha))$ -competitive oblivious routing. Since by 5.2 it must be that $\alpha \geq \text{bal}(G)$, this can be obtained easily by taking the maximum spanning in- and out-arborescences from any fixed vertex.

If we run $\text{ALMOST-ROUTE-DIRECTED}$ with $f_0 = \vec{0}$, we can find $(1 + \epsilon)$ -approximate solutions in time $\tilde{\mathcal{O}}(m\alpha^2/\epsilon^3)$. In order to improve the dependency on ϵ , we can employ a general form of composite

function minimization, introduced in Section 5.4. Define

$$\psi(f) \stackrel{\text{def}}{=} \begin{cases} \infty & \text{if for some } e, f_e/w(e) \notin [0, 50 \ln n / \epsilon] \\ \|\mathbf{C}^{-1}f\|_\infty & \text{otherwise.} \end{cases}$$

The faster algorithm is presented in Figure 7.

$(f, v) = \text{FAST-ALMOST-ROUTE}(b, \epsilon)$

1. Set f_0 using $\text{ALMOST-ROUTE-DIRECTED}(b, \frac{1}{2}, \vec{0})$, keeping the rescaling.
2. Set $K := \lceil \alpha^2 / \epsilon^2 \rceil$.
3. For $k = 0, \dots, K - 1$ let

$$f_{k+1} := \operatorname{argmin}_{f \in \mathbb{R}^E} \left(\nabla \mu(f_k)^\top f + \frac{\alpha^2}{2} \|\mathbf{C}^{-1}(f - f_k)\|_\infty^2 + \psi(f) \right).$$
4. Return $\text{ALMOST-ROUTE-DIRECTED}(b, \epsilon, f_K)$.

Figure 7: Faster algorithm for computing the maximum flow and minimum congested cut.

If we apply the analysis from Section 5.4 (encapsulated in Theorem 5.9), we obtain the following.

Lemma 5.7 *FAST-ALMOST-ROUTE(b, ϵ) terminates in $\tilde{O}(m\alpha^2/\epsilon^2)$ time, assuming $\epsilon \leq 1/2$.*

5.2 Computing Imbalance

As verifying balance can be reduced to a maximum flow computation by Theorem 2.3, we obtain the following result:

Lemma 2.9 *There is an algorithm that either certifies that $\text{bal}(G) \leq \alpha$ or shows that $\text{bal}(G) > (1 - \epsilon)\alpha$ in time $\tilde{O}(m\alpha^2/\epsilon^2)$.*

5.3 Application to Directed Sparsest Cut

In this subsection, assume $G = (V, E)$ is a directed graph that is unweighted, strongly connected, simple, and with an even number of vertices. We define the sparsity of a cut $(S, V \setminus S)$ as $\frac{w(S, V \setminus S)}{|S| \cdot |V \setminus S|}$, where $w(S, V \setminus S)$ is the number of edges going from S to $V \setminus S$. Note that under this definition, no cut can have sparsity greater than one.

As a second application of our maximum flow algorithm, we get the following sparsest cut algorithm. While blocking flows could also possibly be used for our purpose, our approach is clean and may easily generalize to weighted graphs. We defer the details to the appendix.

Lemma 5.8 *Given $\phi \leq 1$, we can find a cut of sparsity ϕ in G or determine that all cuts in G have sparsity $\Omega(\phi/\log^2 n)$ in time $\tilde{O}(m/\phi^2)$.*

5.4 Composite Function Minimization

In this section, we provide a non-Euclidean gradient descent method for minimizing a composite function $f(x) \stackrel{\text{def}}{=} g(x) + \psi(x)$, where g and ψ have specific properties. The algorithm and its convergence guarantee are encapsulated in the following theorem, and they build on several works in convex optimization, such as [Nes13, RT14].

Theorem 5.9 *Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a convex function given by $f(x) \stackrel{\text{def}}{=} g(x) + \psi(x)$ where g is*

convex and L -smooth³ with respect to some norm $\|\cdot\|$. Moreover, assume that $f(x)$ is only finite on some region of diameter D in $\|\cdot\|$. Starting with some $x_0 \in \mathbb{R}^n$ for all k let

$$x_{k+1} := \operatorname{argmin}_{x \in \mathbb{R}^n} \left(\langle \nabla g(x_k), x \rangle + \frac{L}{2} \|x - x_k\|^2 + \psi(x) \right).$$

Then for all $k \geq 1$ we have

$$\epsilon_k \leq \max \left\{ \frac{2 \cdot L \cdot D^2}{\lfloor \frac{k-1}{2} \rfloor + 4}, \left(\frac{1}{2} \right)^{\lfloor \frac{k-1}{2} \rfloor} \epsilon_0 \right\}$$

where $\epsilon_k = f(x_k) - \min_x f(x)$.

Note that the norm we use is arbitrary and we get a gradient descent analysis without appealing to the dual norm. Also we do not require convex ψ we only require convex f .

Acknowledgments

We thank Yin-Tat Lee for several helpful discussions, and in particular for his help with the results in Section 5. This work was partially supported by NSF awards 0843915, 1065106 and 1111109, NSF Graduate Research Fellowship (grant no. 1122374) and Sansom Graduate Fellowship in Computer Science. Part of this work was done while authors were visiting the Simons Institute for the Theory of Computing, UC Berkeley.

References

- [ACF⁺03] Yossi Azar, Edith Cohen, Amos Fiat, Haim Kaplan, and Harald Räcke. Optimal oblivious routing in polynomial time. In Lawrence L. Larmore and Michel X. Goemans, editors, *Proc. of ACM STOC*, pages 383–388. ACM, 2003. 1
- [AKPW95] N. Alon, R. Karp, D. Peleg, and D. West. A graph-theoretic game and its application to the k -server problem. *SIAM J. Comput.*, 24(1):78–100, 1995. 2.3
- [Awe85] Baruch Awerbuch. Complexity of network synchronization. *J. ACM*, 32(4):804–823, October 1985. 2.3
- [Bar96] Y. Bartal. Probabilistic approximation of metric spaces and its algorithmic applications. In *Foundations of Computer Science, 1996. Proceedings., 37th Annual Symposium on*, pages 184–193, 1996. 2.3, 3.2
- [BKR03] Marcin Bienkowski, Mirosław Korzeniowski, and Harald Räcke. A practical algorithm for constructing oblivious routing schemes. In *SPAA 2003: Proceedings of the Fifteenth Annual ACM Symposium on Parallelism in Algorithms and Architectures, June 7-9, 2003, San Diego, California, USA (part of FCRC 2003)*, pages 24–33. ACM, 2003. 1
- [BL99] Yair Bartal and Stefano Leonardi. On-line routing in all-optical networks. *Theoretical Computer Science*, 221(1-2):19–39, 1999. 1, 1.1
- [CKM⁺14] Michael B. Cohen, Rasmus Kyng, Gary L. Miller, Jakub W. Pachocki, Richard Peng, Anup Rao, and Shen Chen Xu. Solving sdd linear systems in nearly $m \log^{1/2} n$ time. In *STOC*, pages 343–352, 2014. C
- [CMP⁺14] Michael B. Cohen, Gary L. Miller, Jakub W. Pachocki, Richard Peng, and Shen Chen Xu. Stretching stretch. *CoRR*, abs/1401.2454, 2014. C

³A function is L -smooth with respect to the norm $\|\cdot\|$ if, for all \vec{x} and \vec{y} , $\|\nabla f(\vec{x}) - \nabla f(\vec{y})\| \leq L\|\vec{x} - \vec{y}\|$.

-
- [ER09] Matthias Englert and Harald Räcke. Oblivious routing for the lp-norm. In *Proc. of IEEE FOCS*, pages 32–40. IEEE Computer Society, 2009. [1](#)
 - [GR98] Andrew V. Goldberg and Satish Rao. Beyond the flow decomposition barrier. *J. ACM*, 45(5):783–797, 1998. [1](#), [1.1](#)
 - [HKLR05] Mohammad Taghi Hajiaghayi, Jeong Han Kim, Tom Leighton, and Harald Räcke. Oblivious routing in directed graphs with random demands. In Harold N. Gabow and Ronald Fagin, editors, *Proceedings of the 37th Annual ACM Symposium on Theory of Computing, Baltimore, MD, USA, May 22-24, 2005*, pages 193–201. ACM, 2005. [1](#)
 - [HKLR06] Mohammad Taghi Hajiaghayi, Robert D. Kleinberg, Frank Thomson Leighton, and Harald Räcke. New lower bounds for oblivious routing in undirected graphs. In *Proc. of ACM-SIAM SODA*, pages 918–927. ACM Press, 2006. [1](#)
 - [HKRL07] Mohammad Taghi Hajiaghayi, Robert D. Kleinberg, Harald Räcke, and Tom Leighton. Oblivious routing on node-capacitated and directed graphs. *ACM Transactions on Algorithms*, 3(4), 2007. ([document](#)), [1](#), [1.1](#)
 - [KLOS14] Jonathan A. Kelner, Yin Tat Lee, Lorenzo Orecchia, and Aaron Sidford. An almost-linear-time algorithm for approximate max flow in undirected graphs, and its multi-commodity generalizations. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, Oregon, USA, January 5-7, 2014*, pages 217–226, 2014. [1](#), [1.1](#), [5.1](#)
 - [Lou10] Anand Louis. Cut-matching games on directed graphs. *CoRR*, abs/1010.1047, 2010. [F](#), [4](#)
 - [LS13] Yin Tat Lee and Aaron Sidford. Matching the universal barrier without paying the costs : Solving linear programs with $\tilde{O}(\sqrt{\text{rank}})$ linear system solves. *CoRR*, abs/1312.6677, 2013. [1](#), [1.1](#)
 - [Mad10] Aleksander Madry. Fast approximation algorithms for cut-based problems in undirected graphs. In *FOCS*, pages 245–254. IEEE Computer Society, 2010. [5.1](#)
 - [Mad13] Aleksander Madry. Navigating central path with electrical flows: From flows to matchings, and back. In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013, 26-29 October, 2013, Berkeley, CA, USA*, pages 253–262, 2013. [1](#), [1.1](#)
 - [MMVW97] Bruce M. Maggs, Friedhelm Meyer auf der Heide, Berthold Vöcking, and Matthias Westermann. Exploiting locality for data management in systems of limited bandwidth. In *Proc. of IEEE FOCS*, pages 284–293. IEEE Computer Society, 1997. [1](#), [1.1](#)
 - [MPX13] Gary L. Miller, Richard Peng, and Shen Chen Xu. Parallel graph decompositions using random shifts. In *Proceedings of the Twenty-fifth Annual ACM Symposium on Parallelism in Algorithms and Architectures, SPAA '13*, pages 196–203, New York, NY, USA, 2013. ACM. [2.3](#), [3.1](#), [3.1](#), [2](#)
 - [Nes13] Yurii Nesterov. Gradient methods for minimizing composite functions. *Math. Program.*, 140(1):125–161, 2013. [5.4](#)
 - [Pen14] Richard Peng. A note on cut-approximators and approximating undirected max flows. *CoRR*, abs/1411.7631, 2014. [1](#), [1.1](#), [5.1](#)
 - [Räc02] Harald Räcke. Minimizing congestion in general networks. In *Proc. of IEEE FOCS*, pages 43–52. IEEE Computer Society, 2002. [1](#)
 - [Räc08] Harald Räcke. Optimal hierarchical decompositions for congestion minimization in

-
- networks. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing, Victoria, British Columbia, Canada, May 17-20, 2008*, pages 255–264, 2008. [1](#), [1.1](#), [2.3](#), [3.2](#), [3.3](#), [3](#), [3.4](#), [D](#)
- [Räc09] Harald Räcke. Survey on oblivious routing strategies. In Klaus Ambos-Spies, Benedikt Löwe, and Wolfgang Merkle, editors, *Mathematical Theory and Computational Practice, 5th Conference on Computability in Europe, CiE 2009, Heidelberg, Germany, July 19-24, 2009. Proceedings*, volume 5635 of *Lecture Notes in Computer Science*, pages 419–429. Springer, 2009. [1.1](#)
- [RST14] Harald Räcke, Chintan Shah, and Hanjo Täubig. Computing cut-based hierarchical decompositions in almost linear time. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, Oregon, USA, January 5-7, 2014*, pages 227–238, 2014. [1](#)
- [RT14] Peter Richtárik and Martin Takác. Iteration complexity of randomized block-coordinate descent methods for minimizing a composite function. *Math. Program.*, 144(1-2):1–38, 2014. [5.4](#)
- [She13] Jonah Sherman. Nearly maximum flows in nearly linear time. In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013, 26-29 October, 2013, Berkeley, CA, USA*, pages 263–269, 2013. [1](#), [1.1](#), [5.1](#), [5.1](#), [5.1](#)

A Missing proofs from Section 2

Lemma A.1 *Let G be a strongly connected directed graph. If demand vector d can be routed in G with congestion c , then $-d$ can be routed in G with congestion at most $\text{bal}(G) \cdot c$.*

Proof:

Note that for any $v \in \mathbb{R}^n$

$$\|\mathbf{U} \max(\mathbf{B}v, 0)\|_1 \leq \text{bal}(G) \|\mathbf{U} \max(-\mathbf{B}v, 0)\|_1$$

follows from the definition of balance.

Hence it is easily seen that the optimum value for the dual problem is within a factor $\text{bal}(G)$ for demands d and $-d$. Our theorem now follows from strong duality to the original problem. ■

Lemma A.2 *Let $l, r \in \mathbb{R}$ with $l \leq r$. Let $C \subseteq \mathbb{R}^m$ be a convex set such that for any $S \subseteq \{1, 2, \dots, m\}$ there exists a point $x \in C$ such that x_i is at least l for $i \in S$ and x_i is at most r for $i \notin S$. Then there exists a point in C with all coordinates in $[l, r]$.*

Proof: Let $P_i(S)$, for $i \in \{0, \dots, m\}$, $S \subseteq \{1, \dots, m\}$ be the subset of points $x \in C$ that satisfy

- $x_j \in [l, r]$ for $j \leq i$ and
- $x_j \geq l$ for $j \in S$ and
- $x_j \leq r$ for $j \notin S$.

We prove that $P_i(S)$ is nonempty for every i and S by induction on i . The base case $i = 0$ follows from the assumption on C . Assume $i \in \{0, \dots, m-1\}$ and the thesis holds for i . Let S be any subset of $\{1, \dots, m\}$. Let $S_L := S \cup \{i+1\}$, $S_R = S \setminus \{i+1\}$. Pick any $x_L \in P_i(S_L)$ and $x_R \in P_i(S_R)$. Then a convex combination of x_L and x_R must belong to $P_{i+1}(S)$. Since S was arbitrary, this concludes the proof. ■

Proof of Theorem 2.3: The implication (2. \rightarrow 1.) and the equivalence (2. \leftrightarrow 3.) are easy to check (note that the circulation of 2. is the sum of $\vec{1}$ and a routing of $-d$). We now prove that if $\text{bal}(G) \leq \alpha$ there exists a circulation in G with each congestion in $[1, \alpha]$.

Note that for any subset S of edges of G we can route the residual degree d_S induced by these edges with congestion 1. Hence by Lemma A.1 we can route $-d_S$ with congestion at most α . Adding these flows yields a circulation with congestion in $[1, \alpha + 1]$ on edges in S and in $[0, \alpha]$ on the other edges. Since the choice of S was arbitrary, the thesis follows by Lemma A.2. ■

We now prove the following lemma, implying 2.2.

Lemma A.3 *Let $G = (V, E, w)$ be an undirected graph and $s, t \in V$. Let d be a demand vector that can be routed in G with congestion at most 1. Let f be a flow from s to t in G with congestion not exceeding 1 satisfying demands $(1 - \epsilon)d$. Let H be the residual graph of f in G . Then $\text{bal}(H) \leq (2\epsilon^{-1} - 1)$.*

Proof: We use the third equivalent definition of balance from Theorem 2.3. The residual degrees in H are $2(\epsilon - 1)d$. Since there exists a flow satisfying demands ϵd with congestion 1, $2(1 - \epsilon)d$ can be routed in H with congestion $2(1 - \epsilon)\epsilon^{-1} = 2\epsilon^{-1} - 2$. ■

B Missing proofs from Section 3.1

Before we prove Lemma 3.1, we shall study the properties of a class of two-way infinite sequences.

For better understanding, we now attempt to provide the intuition on how the sequences defined below are used in the proof. For simplicity, assume we are trying to analyze the clustering of a *path* rather than a cycle. Imagine that every vertex v in the graph sends a runner of unit speed to every vertex of the path, starting at time $-x_v$. After reaching the path, the runner keeps running on it until they reach the end. We will call a runner a *local leader* if they were the first one to reach the end of the path out of all the runners that entered the path at a no later position. It is easy to see that the sequence of origins of local leaders in the order they reach the end of the path is the same as the sequence of roots of clusters into which the path is partitioned by the algorithm. Therefore, it is enough to observe the local leaders as they reach the end of the path. It can be shown that in any time interval $[y, y + \epsilon]$ the probability of the origin of the last local leader to reach the end of the path changing is $\mathcal{O}(\beta\epsilon)$. Unfortunately, the entire process could take an arbitrary amount of time in the case of a path.

To apply the above reasoning to a cycle, we will 'unroll' it into a two-way infinite path. We will set the 'finish line' at an arbitrary vertex (at position 0) and observe the local leaders for any period of time $[y, y + L]$.

Assume the length of the cycle is L and it has l vertices. Let $i \in \{0, \dots, l - 1\}, i \in \mathbb{Z}$. Then, the $i \cdot n + j$ -th element of the sequence s will intuitively be equal to the time the runner sent from the j -th vertex of the graph to the i -th vertex of the unrolled cycle reaches vertex 0 of the unrolled cycle. The sequence a will simply label the origin of the runner relevant to the current index (and so $a_i = (i \bmod n)$). The sequence c will label the cluster to which the relevant vertex of the cycle is assigned (the origin of the first runner to reach it). The function $f(y)$ will give the origin of the runner that reached vertex 0 before time y and entered the cycle at the earliest position. Since only such runners will correspond to clusters, our goal will be to bound the frequency with which f may change.

B.1 Periodically Decreasing Sequences

Let $k, n \in \mathbb{N}$ and $L \in \mathbb{R}_+$.

Let s_i be a two-way infinite sequence of real numbers indexed by $i \in \mathbb{Z}$ with the property

$$\forall_i s_{i+k} = s_i - L.$$

Let a_i be a two-way infinite sequence of integers in $\{0, \dots, n-1\}$ indexed by $i \in \mathbb{Z}$, periodic with period k , that is

$$\forall_i a_{i+k} = a_i.$$

We construct the sequence c_i by defining

$$c_i = a_j,$$

where j is the minimum q that minimizes the value of s_q among $q \leq i$.

We can similarly construct $f : \mathbb{R} \rightarrow \{0, \dots, n-1\}$ by setting for every real number y

$$f(y) = a_j,$$

where j is the minimum q that satisfies $s_q \leq y$.

Fact B.1 *The sequence c_i is periodic with period k .*

Fact B.2 *The function f is periodic with period L .*

Fact B.3 *For any $i \in \mathbb{Z}$ and $y \in \mathbb{R}$, the number of times the sequence $c_i, c_{i+1}, \dots, c_{i+k}$ changes values is equal to the number of times f changes values on the interval $[y, y+L]$.*

B.2 Random Periodically Decreasing Sequences

Let $k, n \in \mathbb{N}, L \in \mathbb{R}_+$.

Let t_i be a two-way infinite sequence of real numbers indexed by $i \in \mathbb{Z}$ with the property

$$\forall_i t_{i+k} = t_i - L.$$

Let a_i be a two-way infinite sequence of integers in $\{0, \dots, n-1\}$ indexed by $i \in \mathbb{Z}$, periodic with period k , that is

$$\forall_i a_{i+k} = a_i.$$

Let x_0, x_1, \dots, x_{n-1} be independent random variables drawn from the exponential distribution with parameter β .

We define for every $i \in \mathbb{Z}$:

$$s_i = t_i - x_{a_i}$$

We define the function $f : \mathbb{R} \rightarrow \{0, \dots, n-1\}$ as in the previous section, that is

$$f(y) = a_j,$$

where j is the minimum q that satisfies $s_q \leq y$.

In the following lemmas, our goal will be to bound the expected number of times the value of f changes on any interval.

Lemma B.4 *For any $y \in \mathbb{R}, \epsilon \in \mathbb{R}_+$, the probability that f is not constant on the interval $[y, y + \epsilon]$ is bounded by $\mathcal{O}(\beta\epsilon)$.*

Proof: Fix y and ϵ . We condition on the value of $f(y + \epsilon)$; assume it is k . We also condition on x_i for all $i \neq k$. Now the condition $f(y + \epsilon) = k$ is equivalent to assuming $x_k \geq c$ for some constant c . Because we have no more information about x_k , the conditional probability that $x_k \geq c + \epsilon$ is $1 - \mathcal{O}(\beta\epsilon)$. This implies the thesis. ■

In order to exploit Lemma B.4 to bound the expected number of changes in f we will attempt to condition on the event D_ϵ .

Definition B.5 *Let $\epsilon \in \mathbb{R}_+$. The event D_ϵ occurs iff for all pairs $i, j \in \mathbb{Z}$ such that $a_i \neq a_j$ or $t_i \neq t_j$ it holds that*

$$|s_i - s_j| > \epsilon.$$

Fact B.6

$$\lim_{\epsilon \rightarrow 0} P(D_\epsilon) = 1.$$

Using B.6, we pick an $\epsilon > 0$ that satisfies

$$P(D_\epsilon) \geq 1 - \min\left(\frac{1}{2}, \frac{\beta L}{k}\right) \quad \text{and} \quad \frac{L}{\epsilon} \in \mathbb{N}.$$

Lemma B.7 *Assume ϵ is chosen as above. Conditioning on D_ϵ , for any $y \in \mathbb{R}$, the probability that f is not constant on the interval $[y, y + \epsilon]$ is bounded by $\mathcal{O}(\beta\epsilon)$.*

Proof: Because $P(D_\epsilon) \geq \frac{1}{2}$, the conditional probability is at most two times larger than in the case where we do not condition on D_ϵ . The thesis follows from Lemma B.4. ■

Lemma B.8 *Assume ϵ is chosen as above. Conditioning on D_ϵ , for any $y \in \mathbb{R}$, the expected number of times f changes values in $[y, y + L]$ is bounded by $\mathcal{O}(\beta L)$.*

Proof: Because we assume D_ϵ , we know that f can change at most once on any interval of length ϵ . Hence it follows from Lemma B.7 that the expected number of time f changes on any interval of length ϵ is bounded by $\mathcal{O}(\beta\epsilon)$. Because $L/\epsilon \in \mathbb{N}$, we can cover the interval $[y, y + L]$ with L/ϵ intervals of length ϵ . Because of linearity of expectation, the expected number of times f changes values on $[y, y + L]$ is therefore bounded by $\mathcal{O}(\beta L)$. ■

Lemma B.9 *For any $y \in \mathbb{R}$, the expected number of times f changes values in $[y, y + L]$ is bounded by $\mathcal{O}(\beta L)$.*

Proof: It follows from B.3 that f cannot change values more than k times on an interval of length L . For ϵ chosen as above, we can apply this observation together with Lemma B.8 to see

that the expected number of changes is bounded by

$$P(D_\epsilon)\mathcal{O}(\beta L) + (1 - P(D_\epsilon))k = \mathcal{O}(\beta L) + \frac{\beta L}{k} \cdot k = \mathcal{O}(\beta L)$$

■

B.3 Low-radius Decompositions

Recall that we are considering the clustering algorithm CLUSTER-DIRECTED applied to a directed graph $G = (V, E)$. Consider a cycle C in G . Assume the length of C is L and the number of vertices on C is l . Let the vertices on the cycle be u_0, \dots, u_{l-1} , in order, with u_0 chosen arbitrarily. For $i \in \{0, \dots, l-1\}$, define p_i to be the distance from u_0 to u_i when going along the cycle.

Let $k = l \cdot n$. We now define the two-way infinite sequence t as follows, for $z \in \mathbb{Z}, i \in \{0, \dots, m-1\}, j \in \{0, \dots, n-1\}$:

$$t_{z \cdot k + i \cdot n + j} = d(v_j, u_i) - zL - p_i.$$

We define the two-way infinite sequence a for $z \in \mathbb{Z}, j \in \{0, \dots, n-1\}$:

$$a_{z \cdot n + j} = j.$$

Fact B.10 *Let $i \in \{0, \dots, l-1\}$. Assume j is a (possibly negative) integer such that $j \leq i \cdot n + n - 1$. Then there exists a path from v_{a_j} to u_i of length $t_j + p_i$.*

Fact B.11 *Let $i \in \{0, \dots, l-1\}, q \in \{0, \dots, n-1\}$. There exists an integer $j \leq i \cdot n + n - 1$ such that $a_j = q$ and*

$$t_j + p_i = d(v_q, u_i).$$

Recall that in CLUSTER-DIRECTED we associate with each vertex v_i an independent random variable x_i drawn from the exponential distribution with parameter β . We now define the two-way infinite sequence s as

$$s_i = t_i - x_{a_i}.$$

As in Section B.1 we construct the sequence c_i by defining

$$c_i = a_j,$$

where j is the minimum q that minimizes the value of s_q among $q \leq i$.

Lemma B.12 *For $i \in \{0, \dots, l-1\}$, $c_{i \cdot n + n - 1}$ is the index of the vertex to whose cluster u_i is assigned by CLUSTER-DIRECTED.*

Proof: This follows from Facts B.10 and B.11. ■

We are now ready to prove the main theorem.

Proof of Lemma 3.1: By Lemma B.12, it is enough to bound the number of times c_0, \dots, c_k changes values. By B.3 this reduces to bounding the number of times the associated function $f : \mathbb{R} \rightarrow \{0, \dots, n-1\}$ changes values on any interval of length L . This is shown to be $\mathcal{O}(\beta L)$ in expectation in Lemma B.9. ■

Proof of Theorem 2.5: First note that with high probability $\max(x_1, \dots, x_n) \leq r$, and so the radius of the computed clusters is at most r . To bound the number of cut edges, it is enough to note that by Theorem 2.3, we can find a set of simple cycles C_1, \dots, C_k such that their total volume is at most $\text{bal}(G)\text{vol}(G)$ and has weight at least $w(e)$ on every edge $e \in E$. The thesis follows by applying Lemma 3.1. ■

C Missing proofs from Section 3.2

Definition C.1 We define the additive imbalance $\text{abal}(G)$ of a directed graph G as the minimum ι such that it is possible to add edges of total weight ι to G to make it Eulerian.

In order to make the running time of our algorithm independent of the diameter of the graph, we will attempt to *collapse* very short edges in the upper levels of the recursion, that is, contract their endpoints into a single vertex. This is similar to the scheme proposed in [CMP⁺14, CKM⁺14]. However, this operation is not always feasible in directed graphs; thus, we will only perform the contraction if both endpoints of the edge can reach each other by following only very short edges.

Definition C.2 Let $G = (V, E, w, l)$ be a directed graph and $x_L, x_R \in \mathbb{R}$ be such that $0 < x_L < x_R$. We construct G collapsed to $[x_L, x_R]$ by:

- merging any vertices that can reach each other while following only arcs of length at most x_L , and
- reducing the length of all arcs longer than x_R to x_R .

$T = \text{FIND-ARBORESCENCE}(G, s)$, where $G = (V, E, l)$ is a directed graph and $s \in V$ is such that all vertices in G are reachable from s .

1. If $n = 1$, return a single-vertex graph.
2. Let $r := \max_{v \in V} d_G(s, v)$.
3. Let $r' := r / (c \cdot \log n)$.
4. Let G' be the graph G collapsed to $[r'/n, 2r']$. Let s' be the vertex in G' corresponding to s .
5. Let $V'_1, V'_2, \dots, V'_k := \text{CLUSTER-DIRECTED-ROOTED}(G', s', r')$.
6. Expand the clusters V'_1, \dots, V'_k back into G , obtaining V_1, \dots, V_k .
7. Let G_i be the graph induced by V_i , for $i = 1, \dots, k$, and u_i denote the center of cluster V_i (with $u_1 = s_1$).
8. Let $T' := \bigcup_{i=1}^k \text{FIND-ARBORESCENCE}(G_i, u_i)$.
9. Let T be T' with the arcs (s, u_i) of length $d_G(s, u_i)$ added for each $i = 2, \dots, k$.
10. Return T .

Figure 8: The low-stretch arborescence finding algorithm.

Lemma C.3 Let $G = (V, E, w, l)$, $s \in V$, $r > 0$. Let $V_1, \dots, V_k = \text{CLUSTER-DIRECTED-ROOTED}(G, s, r)$. Then:

- each cluster V_i has radius at most r ,

$(V_1, V_2, \dots) = \text{CLUSTER-DIRECTED-ROOTED}(G, s, r)$, where $G = (V, E, l)$ is a directed graph, $s \in V$ and $r > 0$.

1. Choose r' uniformly at random from $[0, r]$.
2. Let V_1 be the set of vertices at distance at most r' from s .
3. Let G' be the induced graph on $V - V_1$.
4. Let $V_2, V_3, \dots, V_k := \text{CLUSTER-DIRECTED}(G', r)$.
5. Return V_1, V_2, \dots, V_k .

Figure 9: The decomposition algorithm with a specified root.

- the cluster V_1 containing s has radius at most r from s ,
- the expected total weight of edges going between different clusters is $\mathcal{O}(\text{vol}(G) \log n/r + \text{abal}(G) \log n)$, and
- the expected total additive imbalance of the clusters is $\mathcal{O}(\text{vol}(G) \log n/r + \text{abal}(G) \log n)$.

Moreover, the algorithm works in expected linear time.

Proof: First, note that the expected total weight of edges between V_1 and $V - V_1$ is $\mathcal{O}(\text{vol}(G)/r)$. Hence the expected additive imbalances of the cluster on V_1 and that of G' are both $\mathcal{O}(\text{abal}(G) + \text{vol}(G)/r)$.

By the definition of additive imbalance, we can add edges of expected total weight $\mathcal{O}(\text{abal}(G) + \text{vol}(G)/r)$ to G' to make it Eulerian. We obtain the graph G'' by adding such edges, each with length $2r$. The expected volume of G'' is $\mathcal{O}(\text{vol}(G)) + \mathcal{O}(\text{abal}(G) + \text{vol}(G)/r) \cdot 2r = \mathcal{O}(\text{vol}(G) + \text{abal}(G) \cdot r)$. Now by Theorem 2.5 we can partition G'' into clusters of radius at most r , with the expected total weight of edges going between clusters $\mathcal{O}(\text{vol}(G'') \log n/r) = \mathcal{O}(\text{vol}(G) \log n/r + \text{abal}(G) \log n)$. Note that if we remove the added edges, the radii of these clusters cannot change, as the edges have length greater than r ; at the same time, their total additive imbalance can increase by at most $\mathcal{O}(\text{abal}(G) + \text{vol}(G)/r)$ in expectation. To complete the analysis, observe that in fact the edges added in the above reasoning are ignored by the decomposition algorithm. Hence, they are only necessary for the analysis. ■

Lemma C.4 Let $G = (V, E, w, l)$ be a directed Eulerian graph and $x_L, x_R \in \mathbb{R}$ be such that $0 < x_L < x_R$. Let $G' = (V', E', w', l')$ be G collapsed to $[x_L, x_R]$. Then $\text{vol}(G')$ is at most

$$2 \cdot \sum_{e \in E: l(e) > x_L/n} w(e) \min(l(e), x_R).$$

Proof: Since G' is Eulerian, it can be represented as a sum of simple cycles of uniform weight. Consider any such decomposition and take any cycle C in it. Then C must contain an edge of length at least x_L , and it contains at most n edges of length not exceeding x_L/n . Hence, the length of C is at most two times greater than the sum of its edge lengths greater than x_L/n . Summing over all the cycles yields the desired bound. ■

Proof of Theorem 3.3: First, note that by Theorem 2.3 the edge weights in G can be increased to obtain an Eulerian graph with volume at most $\text{bal}(G) \text{vol}(G)$. Since the algorithm is oblivious to weights, it is enough to consider Eulerian graphs in the proof; from now on we assume $\text{bal}(G) = 1$.

Properties 1 and 2 are easy to verify. Assume the constants hidden in the big-oh notation in Lemma C.3 are bounded by c_0 . We set $c := 2c_0 + 4$.

Consider the i -th level (numbering from 0) of the tree of recursive calls of $\text{FIND-ARBORESCENCE}(G, s)$. Let $r_i = r/(c \log n)^i$. It can easily be shown by induction that the radii of the graphs in the i -th level are at most r_i , and the radii of the returned arborescences are at most $2r_i$, since $c \geq 4$. Let ν_i be the total volume of the *collapsed* graphs at level i .

By Lemma C.3 the additive imbalance of the graphs in the i -th level can be bounded by

$$\begin{aligned} & (c_0 \log n)^i \cdot \nu_0 / r_1 \\ & + (c_0 \log n)^{i-1} \cdot \nu_1 / r_2 \\ & + (c_0 \log n)^{i-2} \cdot \nu_2 / r_3 \\ & + \dots \\ & + (c_0 \log n)^1 \cdot \nu_{i-1} / r_i. \end{aligned}$$

Since $c > 2c_0$, the above sum is bounded by

$$(c \log n)^{i+1} \sum_{j < i} (\nu_j / 2^{i-j}).$$

Hence, the total weight of edges cut at level i is at most

$$(c_0 \log n) \left(\nu_i / r_{i+1} + (c \log n)^{i+1} \sum_{j < i} (\nu_j / 2^{i-j}) \right) \leq (c \log n)^{i+2} / 2 \cdot \sum_{j \leq i} (\nu_j / 2^{i-j}).$$

Since the radius of the arborescence returned at level i is at most $2r_i$, we have that the total stretch incurred at level i is at most

$$2r_i \cdot (c \log n)^{i+2} / 2 \cdot \sum_{j \leq i} (\nu_j / 2^{i-j}) \leq (c \log n)^2 \cdot \sum_{j \leq i} (\nu_j / 2^{i-j}).$$

Hence the total stretch is at most

$$\begin{aligned} (c \log n)^2 \cdot \sum_i \sum_{j \leq i} (\nu_j / 2^{i-j}) &= (c \log n)^2 \cdot \sum_j \left(\nu_j 2^j \sum_{i \geq j} 2^{-i} \right) \\ &\leq 2(c \log n)^2 \cdot \sum_j \nu_j. \end{aligned}$$

Observe that all the collapsed graphs at level j are subgraphs of G collapsed to $[r_{j+1}/n, 2r_{j+1}]$. Hence, by Lemma C.4, we have

$$\nu_j \leq 2 \cdot \sum_{e \in E: l(e) > r_{j+1}/n^2} w(e) \min(l(e), 2r_{j+1}).$$

Hence

$$\begin{aligned} \sum_j \nu_j &\leq 2 \cdot \sum_{e \in E} \sum_{j: r_{j+1} < l(e) \cdot n^2} w(e) \min(l(e), 2r_{j+1}) \\ &= \mathcal{O}(\text{vol}(G) \log n / \log \log n). \end{aligned}$$

Combining this with the previous bound yields the thesis. \blacksquare

D Missing proofs from Section 3.3

Proof of Lemma 3.4: It is enough to note that in step 2c) of Figure 3, with high probability, T'_k is a tree with total stretch $\mathcal{O}(\text{bal}(G) \log^3 n / \log \log n)$ in G'_k , where T'_k and G'_k are undirected counterparts of T_k and G_k , respectively. Hence, the analysis of [Räc08] can be applied to complete the proof. \blacksquare

E Missing proofs from Section 4

Proof of Theorem 2.6: Let $k \geq 1$. Let G be a directed graph on the vertex set

$$V = S \cup T \cup \{s\} \cup \{t\}, \text{ where } |S| = |T| = k$$

and edge set

$$\begin{aligned} E &= S \times T \text{ with weight } 1 \\ &\cup S \times \{s\} \text{ with weight } k \\ &\cup \{(s, t)\} \text{ with weight } k \\ &\cup \{t\} \times T \text{ with weight } k. \end{aligned}$$

Assume some oblivious routing \mathcal{A} achieves competitive ratio c on G . Let $u \in S$ and $v \in T$. The optimal congestion for the unit flow from u to v is at most $1/k$, which can be achieved by routing the flow through s and t . Therefore, \mathcal{A} must achieve congestion at most c/k , hence putting at least $1 - c/k$ units of flow on the edge (s, t) .

The optimal congestion for the multicommodity flow with unit demand between every pair in $S \times T$ is clearly at most 1. Simultaneously, by the above argument, \mathcal{A} must put at least $k(k - c)$ flow on the edge (s, t) . Hence we have $c \geq k - c$, implying $c \geq k/2$. As $n = 2k + 2$ we have $c = \Omega(n)$. \blacksquare

Proof of Theorem 2.7: Let $n \geq 2$. Let G be a directed graph on the vertex set

$$V = \{v_1, \dots, v_n\}$$

and edge set

$$\begin{aligned} E &= C_1 \cup C_2, \text{ where} \\ C_1 &= \{(v_1, v_2), (v_2, v_3), \dots, (v_{n-1}, v_n), (v_n, v_1)\} \text{ with weight } 1, \text{ and} \\ C_2 &= \{(v_n, v_{n-1}), (v_{n-1}, v_{n-2}), \dots, (v_2, v_1), (v_1, v_n)\} \text{ with weight } \sqrt{n}. \end{aligned}$$

Note that G is Eulerian. Assume some oblivious routing \mathcal{A} achieves competitive ratio c on G . Let $i < n$. The optimal congestion for the unit flow from v_i to v_{i+1} is at most $1/\sqrt{n}$, which can be

achieved by routing the flow through C_2 . Therefore, \mathcal{A} must achieve congestion at most c/\sqrt{n} , hence putting at least $1 - c/\sqrt{n}$ units of flow on the edge $(v_n, 1)$.

The optimal congestion for the multicommodity flow with unit demand between every such pair (v_i, v_{i+1}) is clearly at most 1. Simultaneously, by the above argument, \mathcal{A} must put at least $(n-1)(1 - c/\sqrt{n})$ flow on the edge $(v_n, 1)$. Hence we have $c \geq (n-1)/\sqrt{n} - c$, implying $2c \geq \sqrt{n-1}$. Therefore $c = \Omega(\sqrt{n})$.

F Missing proofs from Section 5

Proof of Lemma 5.5:

We have

$$\nabla\mu(f) = -2\alpha\mathbf{B}^\top v.$$

Therefore

$$2\alpha \cdot \|\mathbf{C} \max(\mathbf{B}^\top v, 0)\|_1 \leq 1 + \frac{\epsilon}{4}.$$

It also holds that

$$\begin{aligned} \|\mathbf{C}^{-1}f\|_\infty + 2\alpha v^\top (b - \mathbf{B}f) &= \|\mathbf{C}^{-1}f\|_\infty + p^\top x \\ &\geq \phi(f) - 4 \ln n \\ &\geq \left(1 - \frac{\epsilon}{4}\right) \phi(f). \end{aligned}$$

Simultaneously, we have

$$\begin{aligned} \frac{\epsilon}{4} \phi(f) &\geq \frac{\epsilon}{4} \geq \|\mathbf{C}^{-1}f\|_\infty + \nabla\mu(f)^\top f \\ &= \|\mathbf{C}^{-1}f\|_\infty - 2\alpha f^\top \mathbf{B}^\top v. \end{aligned}$$

Hence

$$2\alpha v^\top b \geq \left(1 - \frac{\epsilon}{2}\right) \phi(f),$$

and so

$$\frac{b^\top v}{\|\mathbf{C} \max(\mathbf{B}^\top v, 0)\|_1} \geq \frac{\phi(f)}{1 + \epsilon}.$$

■

Proof of Lemma 5.6: Let us call the iterations between each scaling in step 2a) a *phase*. Since the initial scaling gives us the correct scale to within factor $1 + \epsilon_0$, we will scale at most $\mathcal{O}(\log(1 + \epsilon_0))$ times. Moreover, if $\epsilon_0 < 1/10$, step 2a) will never be executed.

If step 2b) is about to be executed, then

$$\begin{aligned}\phi(f + \delta s) &\leq \phi(f) + \delta + \delta \nabla \mu(f)^\top s + 2\alpha^2 \delta^2 \\ &\leq \phi(f) - \frac{\epsilon \delta}{4} + 2\alpha^2 \delta^2.\end{aligned}$$

If step 2c) is about to be executed, then

$$\begin{aligned}\phi(f - \delta f) &\leq \phi(f) - \delta \|\mathbf{C}^{-1} f\|_\infty - \delta \nabla \mu(f)^\top f + 2\alpha^2 \delta^2 \\ &\leq \phi(f) - \frac{\epsilon \delta}{4} + 2\alpha^2 \delta^2.\end{aligned}$$

In both cases we have

$$\begin{aligned}\frac{\epsilon \delta}{4} - 2\alpha^2 \delta^2 &\geq \frac{\epsilon^2}{40\alpha^2} - \frac{\epsilon^2}{50\alpha^2} \\ &= \frac{\epsilon^2}{200\alpha^2}.\end{aligned}$$

Hence each iteration of steps 2b) and 2c) decreases $\phi(f)$ by at least $\Omega(\epsilon^2 \alpha^{-2})$.

For $\epsilon_0 \geq 1/10$, every scaling in step 2a) increases $\phi(f)$ by at most $\epsilon^{-1} \ln n$. Hence, for such ϵ_0 there can be at most $\tilde{\mathcal{O}}(\log(1 + \epsilon_0) \alpha^2 \epsilon^{-3})$ iterations in total.

For $\epsilon_0 < 1/10$, step 2a) will never be executed. Moreover, the $\phi(f)$ after the initial scaling must be at most $OPT_b + \tilde{\mathcal{O}}(\epsilon_0 \epsilon^{-1})$. Hence steps 2b) and 2c) can be executed at most $\tilde{\mathcal{O}}(\epsilon_0 \alpha^2 \epsilon^{-3}) = \tilde{\mathcal{O}}(\log(1 + \epsilon_0) \alpha^2 \epsilon^{-3})$ times. ■

Proof of Lemma 5.7: As $\phi(\vec{0}) = \tilde{\mathcal{O}}(OPT_b \alpha)$, step 1. works in $\tilde{\mathcal{O}}(m \alpha^2)$ time by Lemma 5.6.

Now note that we can apply Theorem 5.9 to $\psi'(x) + g(x)$ with $\psi'(x) = \psi(\mathbf{C}x)$, $g(x) = \mu(\mathbf{C}x)$, $L = \alpha^2$, $D = 50\epsilon^{-1} \ln n$. This yields that $\phi(f_K) \leq (1 + \tilde{\mathcal{O}}(\epsilon)) OPT_b$. Hence by Lemma 5.6, step 4. runs in $\tilde{\mathcal{O}}(m \alpha^2 \epsilon^{-2})$ time.

The only remaining thing to show is that we can solve the optimization problem in step 3. in $\tilde{\mathcal{O}}(m)$ time. It can be reformulated by introducing an auxiliary variable z :

$$\begin{aligned}\text{minimize}_{f, z} \quad & \nabla \mu(f_k)^\top f + \frac{1}{2} \alpha^2 z^2 + \psi(f) \\ \text{subject to} \quad & \|\mathbf{C}^{-1}(f - f_k)\|_\infty \leq z.\end{aligned}$$

For a fixed z , the problem can easily be solved in $\mathcal{O}(m \log m)$ time by sorting. Hence we can employ ternary search over z to achieve $\tilde{\mathcal{O}}(m)$ runtime. ■

Proof of Lemma 2.9: Construct G' by adding the reverse of G multiplied by $\frac{1}{4\alpha}$ to G . Note that $\text{bal}(G') \leq 4\alpha$. Let b' be the residual degrees in G' . Now by Theorem 2.8 we can compute a 2-overestimate c' to the minimum congestion to route $-b'$ in G' , in time $\tilde{\mathcal{O}}(m \alpha^2)$. Note that we

have

$$\text{bal}(G') \leq c' - 1 \leq 2\text{bal}(G') \leq 2\text{bal}(G).$$

Hence if $c' - 1 > 2\alpha$ we can conclude that $\text{bal}(G) > \alpha$ and return the corresponding cut.

Otherwise, we must have $\text{bal}(G) \leq 2\alpha$. Hence we can compute a $(1 + \epsilon)$ -overestimate c to the minimum congestion to route $-b$ in G , in time $\tilde{O}(m\alpha^2\epsilon^{-2})$, where b are the residual degrees in G . Now we have

$$\text{bal}(G) \leq c - 1 \leq (1 + \epsilon)\text{bal}(G),$$

and so if $c - 1 \leq \alpha$ then $\text{bal}(G) \leq \alpha$, and otherwise we can return a cut proving that $\text{bal}(G) > (1 - \epsilon)\alpha$. ■

Proof of Lemma 5.8: First, we can use Lemma 2.9 to check whether $\text{bal}(G) \leq \phi^{-1}$. If it is not, we can return the smaller direction of the imbalanced cut as the result. Otherwise, we use can apply the cut-matching game algorithm given by Louis [Lou10] for $\phi' = \frac{n\phi}{4}$ ⁴ and reduce the problem to a sequence of $\tilde{O}(1)$ maximum flow queries. Each of the queries fixes some $S \subseteq V$ with $|S| = n/2$ and asks for a flow in G with demands $-\phi'$ on S and ϕ' on $V \setminus S$. We can compute the 2-approximate minimum congestion flow for such a query. If the returned congestion is at most 1, we return the flow. Otherwise, we have a set $T \subseteq V$ which achieves

$$\begin{aligned} b_T/w(T, V \setminus T) &\geq \frac{1}{2}, \\ w(T, V - T) &\leq 2b_T \\ &\leq 2\phi' \min(|T|, |V \setminus T|) \\ &\leq \frac{n}{2}\phi \min(|T|, |V \setminus T|) \\ &\leq \phi|T| \cdot |V \setminus T|. \end{aligned}$$
■

G Missing proofs from Section 5.4

We break the proof into 2 parts, first we prove a lemma about the progress of each gradient step and then we use this to prove the theorem. Let X_* be the set of all optimal solutions to $\min_x f(x)$.

Lemma G.1 (Gradient Descent Progress) *For all $k \geq 0$ we have that for all $x_* \in X_*$*

$$f(x_{k+1}) \leq f(x_k) - \min \left\{ \frac{1}{2L} \left(\frac{\epsilon_k}{\|x_k - x_*\|} \right)^2, \frac{f(x_k) - f(x_*)}{2} \right\}$$

Proof: By the smoothness of g we know that for all $x \in \mathbb{R}^n$ we have

$$f(x) \leq g(x_k) + \langle \nabla g(x_k), x - x_k \rangle + \frac{L}{2} \|x - x_k\|^2 + \psi(x).$$

⁴The rescaling by n is used due to a slightly different definition of sparsity in [Lou10].

By definition of x_{k+1} we then have that

$$f(x_{k+1}) \leq \min_x \left(g(x_k) + \langle \nabla g(x_k), x - x_k \rangle + \frac{L}{2} \|x - x_k\|^2 + \psi(x) \right).$$

Now it follows from the convexity of g that

$$g(x) \geq g(x_k) + \langle \nabla g(x_k), x - x_k \rangle,$$

and combining these yields that

$$f(x_{k+1}) \leq \min_{x \in \mathbb{R}^n} \left(f(x) + \frac{L}{2} \|x - x_k\|^2 \right) \quad (7.1)$$

Since f is convex, for all $\alpha \in [0, 1]$ and $x_* \in X_*$ we have

$$f(\alpha x_* + (1 - \alpha)x_k) \leq \alpha f(x_*) + (1 - \alpha)f(x_k) = f(x_k) - \alpha(f(x_k) - f(x_*)).$$

Consequently

$$\min_{x \in \mathbb{R}^n} \left(f(x) + \frac{L}{2} \|x - x_k\|^2 \right) \leq \min_{\alpha \in [0, 1]} \left(f(x_k) - \alpha(f(x_k) - f(x_*)) + \frac{L\alpha^2}{2} \|x_k - x_*\|^2 \right).$$

By taking the derivative with respect to α of the expression on the right hand side above and setting it to zero, we see that the optimal α satisfies

$$-(f(x_k) - f(x_*)) + \alpha L \|x_k - x_*\|^2 = 0$$

and thus using $\alpha = \min \left\{ \frac{f(x_k) - f(x_*)}{L \|x_k - x_*\|^2}, 1 \right\}$ yields the result, since $f(x_k) - f(x_*) \geq 0$, and when $f(x_k) - f(x_*) \geq L \|x_k - x_*\|^2$, we have

$$\min_{x \in \mathbb{R}^n} \left(f(x) + \frac{L}{2} \|x_k - x\|^2 \right) \leq f(x_*) + \frac{L}{2} \|x_k - x_*\|^2 \leq f(x_*) + \frac{f(x_k) - f(x_*)}{2}.$$

■

Using the lemma, we can complete the proof of the theorem as follows.

Proof of Theorem 5.9: By Lemma G.1 we have that $\epsilon_{k+1} \leq \epsilon_k$ for all k and

$$\epsilon_{k+1} \leq \max \left\{ \epsilon_k - \frac{1}{2L} \left(\frac{\epsilon_k}{D} \right)^2, \frac{\epsilon_k}{2} \right\}$$

Consequently for $k \geq 1$ such that $\epsilon_k - \frac{1}{2L} \left(\frac{\epsilon_k}{D} \right)^2 \geq \frac{\epsilon_k}{2}$ we have

$$\frac{1}{\epsilon_k} - \frac{1}{\epsilon_{k+1}} \leq \frac{\epsilon_{k+1} - \epsilon_k}{\epsilon_k \epsilon_{k+1}} \leq -\frac{1}{2L} \cdot \frac{1}{D^2} \cdot \frac{\epsilon_k}{\epsilon_{k+1}} \leq -\frac{1}{2LD^2}$$

Summing yields that

$$\frac{1}{\epsilon_1} - \frac{1}{\epsilon_k} \leq -\frac{N_k}{2LD^2}$$

where N_k is the number of steps $k \geq 1$ for which $\epsilon_k - \frac{1}{2L} \left(\frac{\epsilon_k}{D}\right)^2 \geq \frac{\epsilon_k}{2}$. Furthermore, clearly by [G.1](#) we have that

$$\epsilon_1 \leq \frac{L}{2} D^2$$

and thus

$$\epsilon_k \leq \frac{2LD^2}{N_k + 4}$$

On the other hand we have that

$$\epsilon_{k+1} \leq \left(\frac{1}{2}\right)^{k-1-N_k}$$

and noting that either $N_k \geq \lfloor \frac{k-1}{2} \rfloor$ or $k-1-N_k \geq \lfloor \frac{k-1}{2} \rfloor$ then yields the result. ■